LAST TIME:

- Chain Rule: $\frac{\partial f}{\partial t_i} = \frac{\partial f}{\partial x_1} \cdot \frac{\partial x_1}{\partial t_i} + \frac{\partial f}{\partial x_2} \cdot \frac{\partial x_2}{\partial t_i} + \dots + \frac{\partial f}{\partial x_n} \cdot \frac{\partial x_n}{\partial t_i}$
- Implicit Function Theorem: If F is differentiable on an open disk $\frac{\partial F}{\partial x_n} \neq 0$, for all p in the disk $F(p) = 0$, then $\frac{\partial x_n}{\partial x_i} = -\frac{\partial F}{\partial x_i} / \frac{\partial F}{\partial x_n}$ (and $x_n = x_n(x_1, \dots, x_{n-1})$ is a function locally at p.

  * It doesn't really need to be $x_n = x_n(x_1, \dots, x_{n-1})$. It could be any of the others as long as the conditions are met.

  Proof $\quad x_n = f(x_1, x_2, \dots, x_{n-1})$

  Apply the Chain Rule to compute $\frac{\partial F}{\partial x_i} = 0 = \frac{\partial F}{\partial x_1} + \frac{\partial x_1}{\partial x_i} + \dots + \frac{\partial F}{\partial x_n} \cdot \frac{\partial x_n}{\partial x_i}$ (using $F(x_1, \dots, x_n) = 0$)

  Unless $i=k$ or $i=n$, $\frac{\partial x_n}{\partial x_i} = 0$, so we see $0 = \frac{\partial F}{\partial x_1} \cdot \frac{\partial x_i}{\partial x_i} + \frac{\partial F}{\partial x_n} \cdot \frac{\partial F}{\partial x_i}$.

$\therefore \frac{\partial x_i}{\partial x_i} = 1$, so $\frac{\partial F}{\partial x_i} = -\frac{\partial F}{\partial x_i} \Big/ \frac{\partial F}{\partial x_n}$ (subtracting $\frac{\partial F}{\partial x_i}$ and dividing by $\frac{\partial F}{\partial x_n}$)

Ex #1: Compute for $x^3+y^3+z^3 = 6xyz+1$, $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.

$x^3+y^3+z^3 = 6xyz+1$ iff $x^3+y^3+z^3-6xyz-1=0$

Use $F(x,y,z) = x^3+y^3+z^3-6xyz-1$

$\frac{\partial F}{\partial x} = 3x^2-6yz$ $\quad \frac{\partial F}{\partial y} = 3y^2-6xz$ $\quad \frac{\partial F}{\partial z} = 3z^2-6xy$

$\therefore$ By IFT: $\frac{\partial z}{\partial x} = -\frac{\partial F}{\partial x} \Big/ \frac{\partial F}{\partial z} = -\frac{3x^2-6yz}{3z^2-6xy}$ and $\frac{\partial z}{\partial y} = -\frac{\partial F}{\partial y} \Big/ \frac{\partial F}{\partial z} = -\frac{3y^2-6xz}{3z^2-6xy}$

GRADIENT AND OPTIMIZATION

GOAL: Extend optimization tricks to functions of several variables.

- The gradient of a function $f(x_1, x_2, \dots, x_n)$ is $\nabla f = \langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \rangle$.
- It can be used to restate lots of our favorite propositions, like:

1. The Chain Rule: $\frac{\partial f}{\partial t_i} = \nabla f \cdot \frac{\partial \vec{x}}{\partial t_i}$ with $\vec{x} = \langle x_1(t_0, \dots, t_k), \dots, x_n(t_0, \dots, t_k) \rangle$.

That's because $\frac{\partial f}{\partial t_i} = \frac{\partial f}{\partial x_1} \cdot \frac{\partial x_1}{\partial t_i} + \frac{\partial f}{\partial x_2} \cdot \frac{\partial x_2}{\partial t_i} + \dots + \frac{\partial f}{\partial x_n} \cdot \frac{\partial x_n}{\partial t_i} = \langle \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \rangle \cdot \langle \frac{\partial x_1}{\partial t_i}, \dots, \frac{\partial x_n}{\partial t_i} \rangle = \nabla f \cdot \frac{\partial \vec{x}}{\partial t_i}$.

2. Directional Derivatives: $D_{\vec{u}} f(\vec{p}) = \nabla f(\vec{p}) \cdot \vec{u}$

That's because if we let $\vec{u}$ be a unit vector and f be a function w/ $\vec{p} \in \operatorname{dom}(f)$, $D_{\vec{u}} f(\vec{p}) = \lim_{h \to 0^+} \frac{f(\vec{p} + h(\vec{u})) - f(\vec{p})}{h}$.

Consider $g(h) = f(\vec{p} + h\vec{u})$. Now, $g(0) = f(\vec{p})$.

Thus, $D_{\vec{u}} f(\vec{p}) = \lim_{h \to 0^+} \frac{g(h) - g(0)}{h} = g'(0)$.

On the other hand, we recognize g as a composition: $g(h) = f(p_1 + hu_1, p_2 + hu_2, \dots, p_n + hu_n)$.

(w/ Chain Rule)

Therefore, $\frac{\partial g}{\partial h} = \nabla f \cdot \frac{\partial \vec{x}}{\partial h} = \nabla f \cdot \langle \frac{d}{dh}[p_1 + hu_1], \frac{d}{dh}[p_2 + hu_2], \dots, \frac{d}{dh}[p_n + hu_n] \rangle = \nabla f \cdot \langle u_1, \dots, u_n \rangle = \nabla f \cdot \vec{u}$

Ex: Compute $D_{\vec{u}} f(\vec{p})$ for $f(x,y) = 4y\sqrt{x}$, $\vec{p} = \langle 4, 1 \rangle$, $\vec{u} = \langle \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \rangle$.

$\nabla f = \langle 2x^{-1/2} y, 4x^{1/2} \rangle$, $\nabla f(\vec{p}) = \langle 2 \cdot \frac{1}{2} \cdot 1, 4 \cdot 2 \rangle = \langle 1, 8 \rangle$

$\therefore D_{\vec{u}} f(\vec{p}) = \nabla f(\vec{p}) \cdot \vec{u} = \langle 1, 8 \rangle \cdot \langle \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \rangle = \frac{1}{\sqrt{2}} - \frac{8}{\sqrt{2}} = -\frac{7}{\sqrt{2}}$

Ex: Compute ∇f for $f(x,y,z) = \frac{xz}{y+z} = xz(y+z)^{-1}$.

$\frac{\partial f}{\partial x} = \frac{z}{y+z}$ $\quad \frac{\partial f}{\partial y} = -\frac{xz}{(y+z)^2}$ $\quad \frac{\partial f}{\partial z} = \frac{(y+z)\frac{\partial}{\partial z}[xz] - xz\frac{\partial}{\partial z}[y+z]}{(y+z)^2} = \frac{x(y+z) - xz}{(y+z)^2} = \frac{xy}{(y+z)^2}$

$\nabla f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \rangle = \langle \frac{z}{y+z}, -\frac{xz}{(y+z)^2}, \frac{xy}{(y+z)^2} \rangle$

Claim: Gradient optimizes directional derivative, i.e. $\nabla f(p)$'s direction $\vec{u} = \frac{\nabla f(p)}{|\nabla f(p)|}$ realizes maximum $D_{\vec{u}} f(\vec{p})$.

(geometric interpretation of $\cdot$ product)

Why? $D_{\vec{u}} f(\vec{p}) = \nabla f(\vec{p}) \cdot \vec{u} = |\nabla f(\vec{p})| |\vec{u}| \cos\theta = |\nabla f(\vec{p})| \cos\theta$ since $\vec{u}$ is a unit vector

To maximize $D_{\vec{u}} f(\vec{p})$, we want to maximize $\cos(\theta) \leq 1$.

The max is attained at $\theta = 0$, so $\vec{u}$ points in the same direction as $\nabla f(\vec{p})$.

Moreover, the max directional derivative is $|\nabla f(p)|$.

Ex: In which direction does $f(x,y,z) = \frac{xz}{y+z}$ attain its max directional derivative at $\vec{p} = \langle 1, 1, -2 \rangle$? What is the max?

$D_{\vec{u}} f(\vec{p})$ is maximized in the direction of $\nabla f(\vec{p})$.

We saw $\nabla f = \langle \frac{z}{y+z}, -\frac{xz}{(y+z)^2}, \frac{xy}{(y+z)^2} \rangle$, so $\nabla f(1,1,-2) = \langle \frac{-2}{1-2}, -\frac{1(-2)}{(1-2)^2}, \frac{1 \cdot 1}{(1-2)^2} \rangle = \langle 2,2,1 \rangle$.

$\therefore D_{\vec{u}} f(\vec{p})$ is max'd at direction $\vec{u} = \frac{1}{3}\langle 2,2,1 \rangle$ with max $|\nabla f(\vec{p})| = 3$.

- Let f be a function. f has:
  1. A local maximum value at $\vec{p}$ when $f(\vec{p}) \geq f(\vec{x})$ for all $\vec{x}$ nearby to $\vec{p}$.
  2. A global maximum value at $\vec{p}$ when $f(\vec{p}) \geq f(\vec{x})$ for all $\vec{x} \in \text{dom}(f)$.

  * We say for either that $\vec{p}$ is a (local/global) maximum point of f.

  3. A (local / global) minimum value defined similarly with just the inequality signs flipped.

Recall: $f(x)=x$ has neither local nor global extrema (maxima/minima). We want to guarantee extrema (if possible).

- A critical point of function f is a point $\vec{p} \in \text{dom}(f)$ s.t. either $\nabla f(\vec{p})$ does not exist or $\nabla f(\vec{p}) = \vec{0}$.
- PROP (Fermat's Extremum Theorem): If f attains a local extremum at $\vec{p}$, then $\vec{p}$ is a critical point of f.